PubH 7445: Homework Assignment 4

October 4, 2019

Due Friday, October 11, 2019

Please hand in a print-out of your answer and R code, and also email your R code to the TA.

1. Suppose that you are planning an experiment with 40 subjects per group. Assuming that the data is normally distributed with standard deviation 2 in both groups, use simulation to approximate the power of a study design that specifies the use of the Wilcoxon rank sum test to test for differences between the 2 groups if the mean in one group is 1 and the mean in the other group is 2. Compare this to using the 2 sample t-test with equal variance.

2. Problem 4.2 from Foulkes text

3. Problem 4.3 from Foulkes text

4. Problem 4.6 from Foulkes text

5. Problem 4.7 from Foulkes text

6. Independently simulate binary indicator variables for 2000 markers for 100 subjects with success probability 0.5. For each subject simulate a normally distributed outcome variable that depends on the first 10 markers with regression coefficients of $10, 9, \ldots, 1$ and has standard deviation 1. For each marker, use a 2 sample t-test with equal variance to test if the marker is associated with the trait. Make a histogram of the resulting set of p-values and compute a q-value for each marker. If we control the FDR using the default options in the `qvalue` function what is the proportion of false positives and true positives that you observe?

7. Simulate 100 independent observations from a normal distribution with mean zero and standard deviation 1 and 50 independent observations from a normal distribution with mean 3 and standard deviation 0.5 (use the command `set.seed(1)` first). Combine the 2 samples into one sample and make a histogram (this is a plot of a *mixture model* distribution). Use the bootstrap to obtain a confidence interval for the standard deviation of this distribution.